1996

# Toward a Unified Genetic Map of Higher Plants, Transcending the Monocot-Dicot Divergence

Andrew H. Paterson
*Texas A & M University - College Station*

Tien-Hung Lan
*Texas A & M University - College Station*

Kim P. Reischmann
*Texas A & M University - College Station*

Charlene Chang
*Texas A & M University - College Station*

Yann-Rong Lin
*Texas A & M University - College Station*

***See next page for additional authors***

**Authors**
Andrew H. Paterson, Tien-Hung Lan, Kim P. Reischmann, Charlene Chang, Yann-Rong Lin, Sin-Chieh Liu, Mark D. Burow, Stanley P. Kowalski, Catherine S. Katsar, Terrye A. DelMonte, Kenneth A. Feldmann, Keith F. Schertz, and Jonathan F. Wendel

# Toward a unified genetic map of higher plants, transcending the monocot–dicot divergence

Sir — Closely related (confamilial) genera often retain large chromosomal tracts in which gene order is co-linear, punctuated by structural mutations such as inversions and translocations[1]. To explore the possibility that conservation of gene order might extrapolate to more distantly related taxa, we first estimated an average structural mutation rate. Nine pairs of taxa, for which there exist both comparative genetic maps and plausible estimates of divergence time, showed an average of 0.14 (±0.06) structural mutations per chromosome per million years of divergence (Myr; Table 1). This value is offered as a first approximation, acknowledging that refined comparative data and/or divergence estimates may impel revision.

A predictive model, based on the structural mutation rate that we estimated, suggests that small regions of common gene order might persist in taxa that are diverged by much longer time periods than those investigated to date (Fig. 1*a*). Data more detailed than those presently available are likely to impel a revised model using different rate constants for the two major types of structural mutations, intra-chromosomal inversions and inter-chromosomal translocations.

The predicted existence of small chromosome segments retaining common gene order after long divergence times, was tested by genetic mapping of common *Arabidopsis* expressed-sequence tags (AESTs) in the flowering plant subclasses *Monocotyledonae* (monocots: *Sorghum* spp.) and *Dicotyledonae* (dicots: *Arabidopsis thaliana*, *Brassica oleracea* [broccoli], *Gossypium* [cotton] spp.). Monocots and dicots collectively include most agricultural crops and botanical models, and diverged from a common ancestor about 130–200 million years ago[2,3]. Over this period, our model (Fig. 1*a*) predicts that 43–58% of chromosomal tracts ≤3 cM should remain co-linear. A comparative map of the crucifers *Arabidopsis* and *Brassica* (ref. 4; T.-H. Lan *et al.*, unpublished data) enabled us to treat these genera essentially as a unit.

Among eight pairs of genes linked at ≤3 cM in the crucifers, seven pairs (87.5%) were also linked in sorghum at distances of 1.4, 13.5, 26.0, 3.3, 9.0, 12.5 and 46.7 cM, respectively (AEST38-51, Fig. 1*b*; AEST8b-239 and 239-69, Fig. 1*c*; AEST56-18a, 18a-39, and 39-146, Fig. 1*d*; AEST136-137, Fig. 1*e*; AEST36-123, Fig. 1*f*). The eighth pair, AEST39-146, mapped to putatively homoeologous sorghum chromosome segments (Fig. 1*d*). Among three pairs of probes linked at ≤3 cM in the crucifers (AEST 51-9 and 9-204, Fig. 1*b*; AEST18a-56, Fig. 1*d*) the first two were linked at 5.8 and 41.1 cM in cotton; the last pair were unlinked. The lesser conservation found in cotton compared with sorghum is presumably an artifact of lower density of comparative markers (average 38 cM spacing, versus 19 cM in sorghum).

Many loci that were more distantly linked in the crucifers remain syntenic (on the same chromosome) in sorghum (Fig. 1*c*, *f*), implying that additional shortest conserved evolutionary unit sequences[5] (SCEUS) may be delimited as more comparative markers are mapped. Large chromosomal intervals are highly likely to incur structural mutation(s) over 200 Myr (Fig. 1*a*). Apparent co-linearity across such intervals cannot be inferred to represent absolute conservation in the order of all intervening genes in monocots and dicots, but may reflect the presence of several smaller gene tracts which

**Table 1 Rates of chromosome structural mutation for different taxa**

| Taxon | Genera (basal chromosome #) | Chromosomal rearrangements | Estimated divergence time (Myr) | Rearrangements per chromosome per Myr |
|---|---|---|---|---|
| *Dicotyledoneae* | | | | |
| *Solanaceae* | *Lycopersicon/Solanum* (12) | 5 (ref. 22) | 10[a] | 0.042 |
| | *Lycopersicon/Capsicum* (12) | 33[b] | 40(±10)[a] | 0.069 |
| *Brassicaceae* | *Arabidopsis/Brassica* (5/9)[c] | 26[4] | <10 (ref. 24) | 0.52 |
| *Malvaceae* | *Gossypium* spp. (13) | 9[d] | 4–11 (ref. 25) | 0.11 |
| *Monocotyledoneae* | | | | |
| *Poaceae* | *Zea/Sorghum* (10) | ≥14 (ref. 12) | 24 (ref. 26) | 0.058 |
| | *Zea/Oryza* (10/12)[c] | 23 (ref. 13) | 66 (ref. 27) | 0.035 |
| | *Oryza/Triticum* (12/7)[c] | ≥20 (ref. 14) | 66 (ref. 27) | 0.043 |
| | *Triticum/Secale* (7/7) | 13 (ref. 23) | 6 (ref. 28) | 0.31 |
| *Mammalia* | | | | |
| Primate/*Rodentia* | *Homo/Mus* (24/20)[c] | 138[e] | 100 | 0.069 |
| Average: | | | | 0.14 (±0.06) |

[a]R. Olmstead, pers. comm. [b]Based upon application of the algorithm described[29] to published data[23]. [c]Rate calculations were based on the smaller chromosome number, so expectations of co-linearity are conservative. [d]Only 11 of 13 homoeologous chromosomes are sufficiently well-mapped to make rearrangements clear[8]. Consequently, the rate calculation used a value of 10.6 rearrangements (9•13/11), and a consensus divergence time of 7.5 Myr. [e]Demonstrated using a map of 241 genes at ~6 cM intervals[30]. Ongoing mapping[5] has defined additional SCEUS, and the final number may approach 180, as previously predicted[29]. This latter value would indicate a somewhat higher rate of rearrangement (0.09) in *Mammalia*, but only nominally affect our overall structural mutation rate estimate.

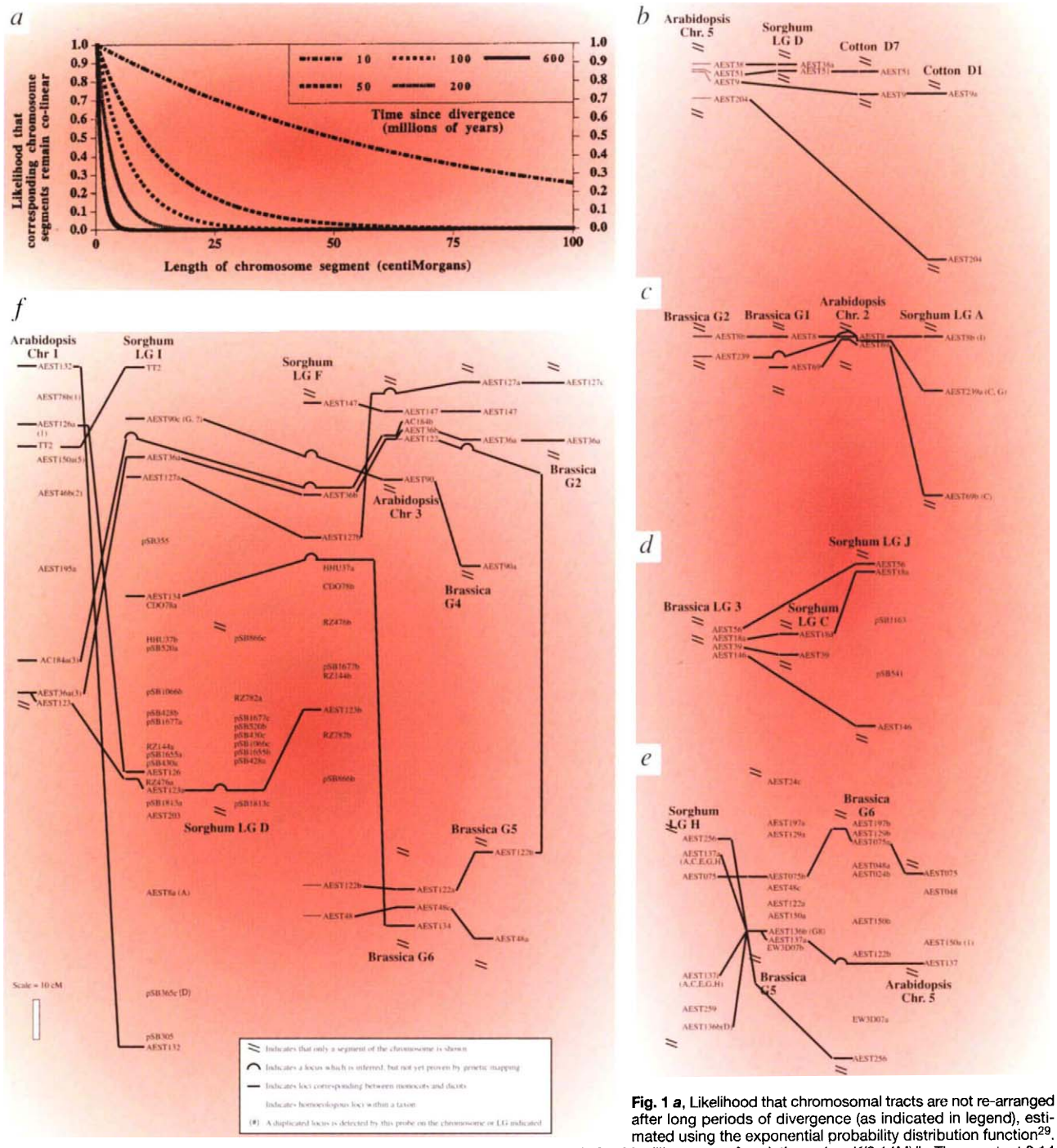


**Fig. 1** ***a***, Likelihood that chromosomal tracts are not re-arranged after long periods of divergence (as indicated in legend), estimated using the exponential probability distribution function[29]. The probability that an interval of K cM, in a chromosome of L cM, has been rearranged after M million years of evolution = 1–e–K(0.14M)/L. The constant 0.14 is the estimated rate of structural mutation, based on an average rate of 9 pairs of taxa (see Table 1). Likelihoods are based on a value of L = 100 cM. ***b–f***, Co-linearity of monocot and dicot genes. *Arabidopsis* cDNAs that show DNA sequence conservation (BLASTx > 150; ref. 31) with genes from monocots or more distant taxa, were used to detect restriction fragment length polymorphisms (RFLPs), and added to existing genetic maps of *Sorghum bicolor* × *S. propinquum*[6], *Arabidopsis thaliana*[4], *Brassica oleracea* (T.-H.L. *et al.*, unpublished data) and *Gossypium trilobum* × *G. raimondii* (C. Brubaker, A.H.P., J.F.W., unpublished data), by described procedures[4,6,8]. *Brassica* mapping followed essentially the same procedures used for *Arabidopsis*, in 56 progeny of a cross between a self-compatible rapid-cycling ***b***, *oleracea* selection (from P. Williams, Madison WI) and the commercial broccoli cultivar 'Green Comet.' AEST, TT, and AC denote *Arabidopsis* cDNAs. pSB denotes sorghum *Pst*I-genomic DNA clones. *HHU*, *CDO*, and *RZ* denote sorghum, oat, and rice cDNAs. Lower-case letters denote that additional loci were detected by the same DNA probe. Chromosomal locations of duplicate loci are shown either directly, or in parentheses. Among 161 probes screened in *Sorghum*, 52 (32%) could be mapped to 79 loci (11 to 2 loci, 4 to 3, 1 each to 4 and 6): 35 of these could be mapped in *Arabidopsis*, and 11 more in *Brassica*. Fewer (randomly chosen) probes were screened in cotton, and 29 mapped to 39 loci (6 to 2 loci, and 2 to 3): 20 of these could be mapped in *Arabidopsis*, and 4 more in *Brassica*. b, A segment of *Arabidopsis* chromosome 5 retains co-linearity with both sorghum and cotton. Co-linearity of duplicated *AEST9* loci with *AEST51* and *AEST204* (respectively) suggests that these segments of cotton linkage groups (LGs) *D7* and *D1* may derive from an ancient duplication. ***c***, Co-linearity of *Arabidopsis* and *Brassica* helps to reveal synteny with *Sorghum* (see *AEST239*). ***d***, *Sorghum* evolution may also have involved chromosomal duplication, consistent with recent data[6,7]. ***e***, Apparent conservation is obscured by the fact that *AEST137* reveals RFLPs at six different sorghum loci; two proximal loci are shown. ***f***, *AEST36*, *AEST123*, and *AEST127* duplicated loci support evidence from sorghum DNA probes that different parts of LG I are homoeologous to parts of LGs F and D. *AEST36* and *AC184* also show closely linked duplicated loci on *Arabidopsis* chromosomes 1 and 3, suggesting that this duplication may predate monocot-dicot divergence. Segments of *Arabidopsis* chromosome 3, and *Brassica* homoeologs G2/G4 and G5/G6 (respectively) are related to different parts of sorghum LG F, and possibly to each other (see *AEST122*). Correspondence of *AEST* numbers to microtiter plate numbers for the clone repository has been deposited at the *Arabidopsis* Biological Resources Center and at the *Nature Genetics* web site (http//genetics.nature.com).

have been reshuffled by intrachromosomal inversion.

The arrangements of duplicated loci detected by these highly conserved DNA probes support recent evidence of ancient chromosomal duplication in sorghum[6,7] (Fig. 1*d*, *f*), cotton[8] (Fig. 1*b*), and even *Arabidopsis* [4,9] (Fig. 1*f*). Conservation of these DNA sequences from before the monocot–dicot divergence should, and apparently does, transcend subsequent chromosomal duplication(s) that have occurred in many plant genomes[10,11]. Pseudogene formation, or other sub-chromosomal duplications, may complicate interpretation of some comparative data (Fig. 1*e*).

Common gene order, as implied by parallel genetic linkage relationships, provides a framework for unifying genetic maps of divergent taxa. Based on a modal SCEUS length of 3 cM, predicted by our model and congruent with our empirical data, as few as 200 rearrangements may distinguish the genomes of *Arabidopsis* and *Sorghum*. Extrapolation of results from sorghum to other grasses[3,6,12–15], and *Arabidopsis* to other crucifers[4], or even to other dicot families such as cotton, will provide a starting point for a unified map. By phylogenetic analysis, ancestral versus derived gene orders might be discerned, revealing the course of chromosome evolution and providing more data to evaluate the need for separate rate constants for inversions and translocations[4,13].

A unified genetic map would afford new opportunities for molecular dissection of both simple and complex phenotypes. Physical maps for facile models such as *Arabidopsis*[16] might aid in the cloning of agriculturally important genes or quantitative trait loci (QTLs)[17] from major crops. Thousands of genetically mapped mutants of *Arabidopsis*, maize, rice, and other taxa might be united into a central tool for comparative study of plant development. Mutants unique to one taxon may facilitate molecular dissection of processes that are invariant in other taxa.

The ability to evaluate phenotypic convergence over long periods of biotic evolution, as already demonstrated across 65 Myr of divergence[18], may have many important consequences. In medicine, comparative analysis might shed new light on convergent or parallel evolution of functionally similar structures, such as the eyes of invertebrates and vertebrates, which evolved independently but may share a common genetic basis[19]. In agriculture, if corresponding genes cause susceptibility of divergent crops to common pests[20], then strategies fundamental to 'integrated pest management' such as 'crop rotation' may require re-evaluation.

Conserved gene blocks that are larger than expected to persist by chance might reflect unusual structural features or genomic processes that confer fitness advantages. Predicted lengths of co-linear chromosome segments (Fig. 1*a*) provide a null hypothesis to identify such properties. Chromosomal regions that are insulated from rearrangement, or taxa that show rapid rearrangement such as *Brassica*[4] may be fruitful systems for new investigations.

Tentative co-linearity in two small regions of the rice and human chromosomes[21] hints at the possibility of a unified map for eukaryotes. Using unified maps as a conduit, opportunities for comparative biology may extend far beyond our present grasp.

**Acknowledgements**

*We thank the Ohio State Univ.–NSF Arabidopsis Biological Resources Center (AEST); S. McCouch, M. Sorrells, S. Tanksley (Cornell Univ.; CDO, RZ); P. Westhoff (Heinrich Heine Univ.; HHU37); T. Thomas (Texas A&M Univ.; TT2) for probes; C. Somerville for lists of conserved Arabidopsis ESTs, J. McFerson & S. Kresovich for Brassica pollen; J. Doebley, J. Irvine, Y-W. Wang, R. Wright, for suggestions; and the Texas Higher Education Coordinating Board, USDA-NRI, Texas, Delaware (A.H.P.), & Arizona (K.A.F.) Agricultural Experiment Stations, NSF (K.A.F., J.F.W.), & USDA-ARS (K.F.S.) for funds.*

**Andrew H. Paterson[1,2]**
**Tien-Hung Lan[1]**
**Kim P. Reischmann[1]**
**Charlene Chang[1]**
**Yann-Rong Lin[1]**
**Sin-Chieh Liu[1]**
**Mark D. Burow[1]**
**Stanley P. Kowalski[1]**
**Catherine S. Katsar[1]**
**Terrye A. DelMonte[1]**
**Kenneth A. Feldmann[3]**
**Keith F. Schertz[4]**
**Jonathan F. Wendel[5]**

*[1]Plant Genome Mapping Laboratory, Department of Soil and Crop Science, Texas A&M University, College Station, Texas 77843, USA*
*[2]Department of Plant and Soil Sciences, University of Delaware, Newark, Delaware 19711, USA*
*[3]Department of Plant Science, University of Arizona, Tucson, Arizona 85721, USA*
*[4]United States Department of Agriculture, Agricultural Research Service, College Station, Texas 77843, USA*
*[5]Department of Botany, Iowa State University, Ames, Iowa 50011, USA*

1. Shields, R. *Nature* **365**, 297–298 (1993).
2. Wolfe, K.H., Gouy, M., Yang, Y., Sharp, P.M. & Li, W.H. *Proc. Natl. Acad. Sci. USA* **86**, 6201–6205 (1989).
3. Crane, P.R., Friis, E.M. & Pedersen, K.R. *Nature* **374**, 27–33 (1995).
4. Kowalski, S.D., Lan, T.-H., Feldmann, K.A. & Paterson, A.H. *Genetics* **138**, 499–510 (1994).
5. O'Brien, S. J. *et al. Nature Genet.* **3**, 103–112 (1993).
6. Chittenden, L.M., Schertz, K.F., Lin, Y.-R., Wing, R.A. & Paterson, A.H. *Theor. Appl. Genet.* **87**, 925–933 (1994).
7. Periera, M.G. *et al. Genome* **37**, 236–243 (1994).
8. Reinisch, A.J. *et al. Genetics*, **138**, 829–847 (1994).
9. McGrath, J.M., Jancso, M.M., & Pichersky, E. *Theor. Appl. Genet.* **86**, 880–888 (1993).
10. Stebbins, G. L. *Science* **152**, 1463–1469 (1966).
11. Masterson, J. *Science* **264**, 421–424 (1994).
12. Whitkus, R., J. Doebley, & M. Lee. *Genetics* **132**, 1119–1130 (1992).
13. Ahn, S. & Tanksley, S.D. *Proc. Natl. Acad. Sci. USA* **90**, 7980–7984 (1993).
14. Kurata, N. *et al. Bio/Technology* **12**, 276–278 (1994).
15. Devos, K., Atkinson, M.D., Chinoy, C.N., Liu, C.J. & Gale, M.D. *Theor. Appl. Genet.* **83**, 931–939 (1992).
16. Schmidt, R. *et al. Science* **270**, 480–483 (1995).
17. Paterson, A.H. *et al. Nature* **335**, 721–726 (1988).
18. Paterson, A.H. *et al. Science* **269**, 1714–1718 (1995).
19. Quiring, R., Walldorf, U., Kloter, U. & Gehring, W.J. *Science* **265**, 785–788 (1994).
20. Marchetti, M. *Plant Dis.* **75**, 773–775 (1991).
21. Kurata, N. *et al. Nature Genet.* **8**, 365–372 (1994).
22. Tanksley, S.D. *et al. Genetics* **132**, 1141–1160 (1992).
23. Prince J.P., Pochard E. & Tanksley S.D. *Genome* **36**, 404–417 (1993).
24. Muller, J. *Bot. Rev.* **47**, 1–142 (1981).
25. Wendel, J. F. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4132–4136 (1989).
26. Thomasson, J.R. Fossil grasses, 1820–1986 and beyond. in *Grass Systematics and Evolution* (eds Soderstrom, T.R., Hilu, K.W., Campbell, C.S. & Barkworth, M.E. ) 159–167 (Smithsonian Institution Press, Washington, 1987).
27. Linder, H.P. *Kew Bull.* **42**, 297–318 (1987).
28. Peterson, G. & Doebley J.F. *Pl. Syst. Evol.* **187**, 115–125 (1993).
29. Nadeau, J.H. & Taylor, B.A. *Proc. Natl. Acad. Sci. USA* **81**, 814–818 (1984).
30. Nadeau, J.H. *et al. Mamm. Genome* **3**, 480–536 (1992).
31. Newman, T. *et al. Plant Physiol.* **106**, 1241–1255 (1994).